Validation of Simulation Models

Robert G. Sargent

Syracuse University

ABSTRACT

This is a tutorial paper on validation of simulation models. Included in this tutorial are what is meant by validation, the problem dependent characteristics of simulation model validation, descriptions of the various validation techniques and their use and a discussion on the statistics used in validation techniques (but not the detailed statistical tests themselves).

INTRODUCTION

Validation is an important step in developing simulation models. Simulation models are developed for specific purposes or objectives. It is against these objectives (or objective) that a model is judged whether it is or is not valid. In validation, one does not prove that the model is valid (correct) under all sets of conditions. Instead, a degree of confidence is obtained that the model and its results are reasonable for the objective it was developed for. It is usually too costly and time consuming to prove a model is valid under all conditions.

Figure one illustrates two relationships of model confidence. One curve illustrates how the validation cost increases as the model confidence increases and the other curve shows the relationship of the value of the model's results to the user as a function of model confidence. Note that the degree of confidence lies between 0% and 100%, where 100% implies that the model has been proven to be correct (valid) for the objectives it was developed for.

Basically three questions are of concern. (1) Does the simulation model behave as a model builder believes? (2) Does the simulation model adequately represent the system for the objective(s) it was developed? (3) Does the simulation model user have confidence in the model's results? Fishman and Kiviat [8] called the process of answering the first question verification and the second question validation. Most authors use this terminology [9, 17, 22]. Others, however, make no distinction between the use of these terms or use them to mean just the opposite, i.e., validation is addressing the first question and verification is addressing the second question. In this paper we will use the terms verification and validation as Fishman and Kiviat do.

Figure two presents a simplified process of the steps of the modelling process of concern in model verification and validation. Question one above is primarily concerned with (a) does the flowchart model represent the system or idea as the modellers (model builders) believe it does, and (b) does the simulator (computerized model) accurately represent the flowchart model? This latter process is sometimes called program correctness. The second question above is concerned with how adequately does the flowchart model and the simulator represent the system or idea under investigation for the specific objectives the simulation model is being developed for. The third question concerns whether the model user believes the model's output or results of interest. (We are assuming in this paper that the computer language is error free, a good random number generator is being used in simulation models, etc. in discussing validation.)

To address these questions, various techniques and approaches are used to obtain confidence in the model. There are currently no known procedures (algorithms) or specific sequences of using the various techniques to develop confidence in simulation models nor are there known ways to quantify the confidence one has in a specific model. Also, many techniques are problem dependent. Thus in practice, the analyst or model builder uses those techniques that seem appropriate to obtain the necessary confidence in the model considering such factors as cost, time, model use, model's users, etc. Furthermore, these questions are usually not addressed entirely separate of each other. Certain techniques and tests are performed for each question separately and others for both or all three questions together.

CH1437-3/79/0497-0503$00.75
1979 Winter Simulation Conference

Validation of Simulation Models (continued)

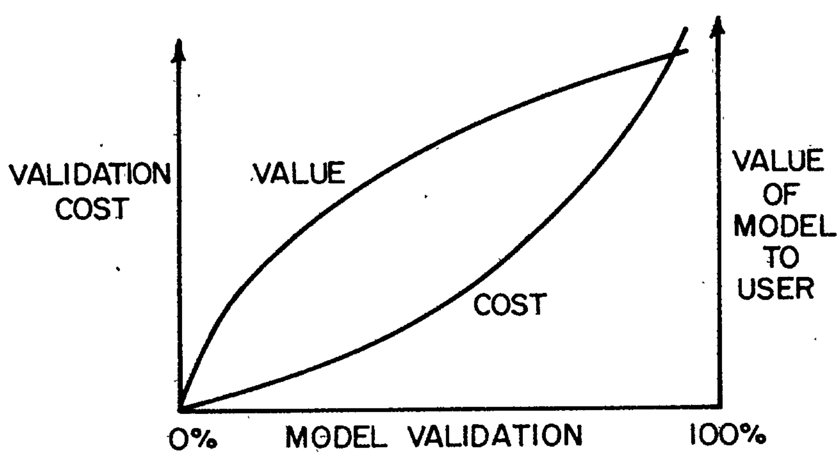
This author believes, as do many others, that model validation is not completely separable from model building. He believes that validation techniques should be used during the model building process and also that model building is an iterative process and confidence in the model is increased from model iteration to model iteration. It is also generally agreed that it is necessary to have system data if reasonable model confidence is to be obtained, preferable for at least two different operating levels.

We can illustrate the relationships of verification and validation to the modelling process more clearly if we consider verification to be comprised of flowchart verification and simulator verification and technical aspects of verification and validation, and the modelling process, then the user must obtain confidence in the model in other ways.

This author believes that a modeller (or an organization) develops "credibility" with model users and as this credibility increases, model users increase their confidence in the models developed by this modeller. One way to develop credibility and user confidence is to have the model user be an active member of the modelling (problem solving) team. Another way is to involve the model user in the application of specific validation techniques, e.g., face validity.

In Figure three, we have included data validity.

FIGURE ONE



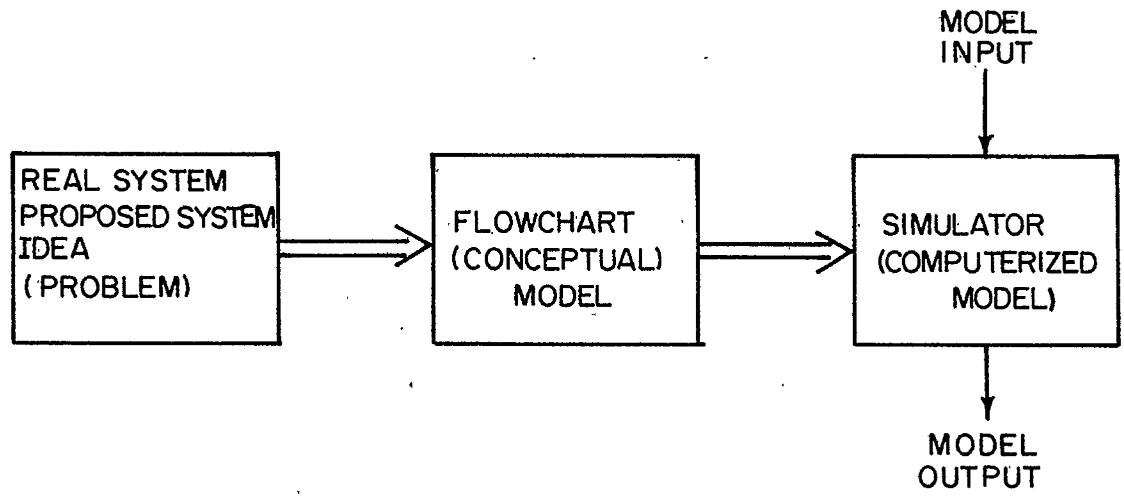
validation to be comprised of theoretical validity and operational validity. Flowchart verification is concerned with whether the flowchart model represents the system (or idea) as the modellers believe it does and simulator verification is concerned with whether the flowchart model has been programmed correctly and that there are no computation problems in the model, e.g., computation errors due not accumulate to make the results inaccurate. Theoretical validity is concerned with whether the theories used and the assumptions made in a model will allow it to have the necessary accuracies required for the problem and operational validity is concerned with the ability of the simulator to reproduce the system or idea such that its results are sufficiently accurate to satisfy the model's objectives or purpose. Using these definitions, we show in Figure three the relationships of verification and validation to the modelling steps given in Figure two.

The user of a simulation model either is or is not able to follow and understand the technical aspect of model building, verification and validation. If the model user is able to follow this process, then the user should obtain confidence in the model through the <u>documentation</u> the modelling team produces. If the user is not able to understand the

Data validity is usually not discussed or mentioned with regard to model validation. However, obtaining sufficient and accurate data for model building and validation is difficult, time consuming, and frequently costly. It has been this author's experience that data validity is extremely important and is the most common reason that initial attempts to validate a model fails. For example, in [2], the model of a rework area of a production system was not able to be validated until a second time study was made of the system to obtain accurate data.

Gass [9, 10] has recently proposed for complex models to have an independent evaluation of each model made by individuals that are neither the developers or users to determine if a user should rely on and use the outputs of a given model. He presents thirteen questions (items) on verification, validation, and documentation for evaluators to use in evaluating models. The evaluator selects a numerical value between zero and ten for each question based upon the analyses and conclusions reached for that question. A numerical rating is obtained for each model by using a set of weights given for the questions and the numerical values selected for each question by the evaluator. The numerical rating obtained for

FIGURE TWO



a model then puts that model into one of three categories: Category I, the model can be used with confidence; Category II, the model is acceptable and useable, but needs further analysis or minor improvements in order to be used with confidence; and Category III, the model should not be used unless major deficiencies are corrected.

Some authors discuss validation criteria other than confidence in the model. Shannon [17] discusses a degree of validity between zero and one which is essentially identical to confidence in the model between zero and one hundred percent. Zeigler [23] states that he believes there are increasing degrees of model validity and defines three such degrees: replicatively valid, the lowest level, which means the model data is able to match previous system data; predictively valid, the next level, which means that the model data can predict future system data; and structurally valid, the highest level, which means the model structure agrees with the internal workings of the real system. Schellenberger [16] recently suggested three basic kinds of criteria for assessing model validity for managerial purposes: technical validity, which is to identify model divergence from the real system; operational validity, which is to identify the effect of the model divergence on the model's use; and dynamic validity, which is to identify the range of the model validity. The U.S. General Accounting Office in its Guidelines for Model Evaluation [11] presents a list of criteria for model evaluation in order to evaluate a model to determine if it should be used in decision making.

The next section of this paper describes various problems characteristics that effect the validation techniques used. This is followed by a section describing the various validation techniques, and a section on using them. The last section is a summary.

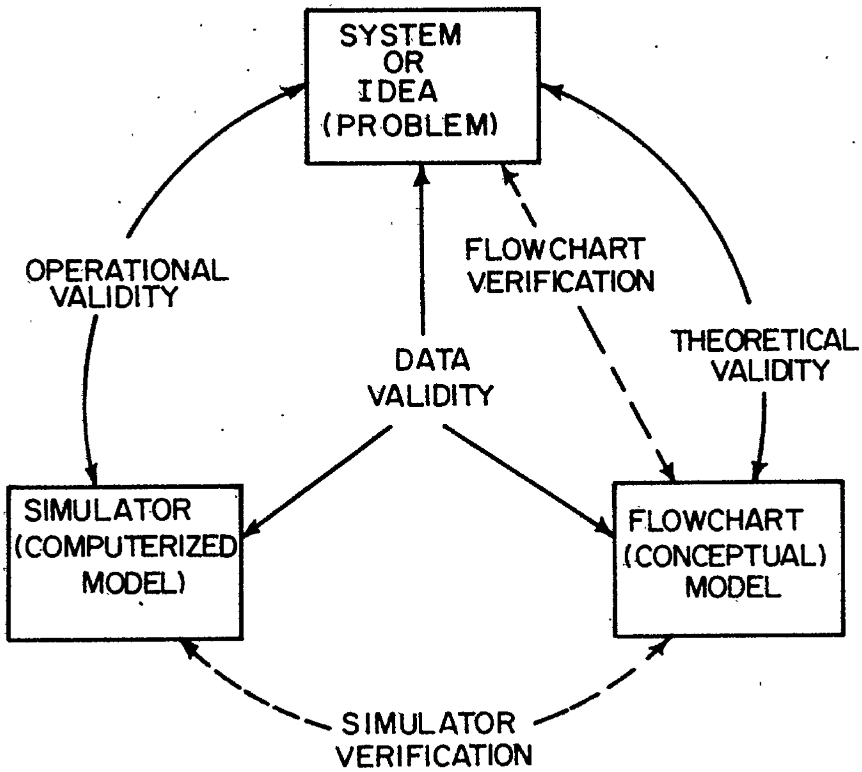
CHARACTERISTICS OF MODELLING PROBLEMS

The objective of this section is to define the various characteristics that effect validation. The characteristics are broken into three separate classes: problem characteristics, system characteristics, and model characteristics.

The class of problem characteristics describes the characteristics of the problem environment and the objectives of developing the model. Is the purpose of the study to investigate system output, system behavior, system mechanisms (structure), or obtain system insights? Is the model to be used in research, development, system design, selection among alternative systems (e.g., new equipment), modification of an existing system (e.g., tuning, peaking), or in an operating system at the strategical, tactical or operational level? Is the model to be used once or many times? Is the model to investigate transient or nonstationary behavior, steady state behavior, or both? What range of system values are of concern? What are the precise purposes and objectives of the model? Do other models of this system exist? What are the cost returns expected from modelling? What is the relationship of the model user to the model developer (close or far removed)? Does the user understand the modelling process, and if so, at what level?

The class of system characteristics characterize the system under investigation. Does the system exist or does it not? Is it a natural system or man-made (artificial) system? If the system exists, can appropriate and sufficient data be obtained on the system and can controlled experiments be made to collect data? (For example, a computer system may be able to run in two different configurations to obtain lots of data whereas only a small amount of data may be able to be obtained on environmental or inventory systems.) How much does it cost to collect data on the system? Is the system deterministic or stochastic? Does the system operate in

FIGURE THREE



steady state or is it a non-stationary system?

The class of model characteristics describe the model itself. What are the assumptions underlying the model? Is the model deterministic or stochastic? Is the model dynamic or static? Does it have submodels? What range of values is it valid over? What type of data collection and analysis is possible? How much does it cost to develop the model? How much does it cost to run the model? What languages is it programmed in? How much documentation is there on the model? What level of model validation is desired?

VALIDATION TECHNIQUES

This section describes the various validation techniques that are used to develop confidence in simulation models. As stated above, a combination of techniques are commonly used in performing a validation. Most of the techniques described here are found in the literature [3,4,5,6,7,8,13,16,17,22], although they may be modified slightly or described differently by this author. Usually it is desirable to use statistical tests with the validation techniques, if appropriate; if not, other techniques should be employed such as graphical plots. The usage of these techniques are discussed in the next section.

Face Validity: Face validity is asking people knowledgeable about the system whether the model is reasonable. One can, for example, apply face validity to the model flowchart to determine if the logic is correct. Also face validity is commonly used to determine if different model input-output relationships and their internal behavior (e.g., queue lengths) are reasonable.

Traces: The behavior of different types of specific entities in the model are traced (followed) through the computerized model to determine if the model's logic and the computer program are correct and if the necessary accuracy is obtained.

Historical Methods: Three historical methods of validation are Rationalism, Empiricism, and Positive Economics. Rationalism assumes everyone knows whether the underlying assumptions of a model are true. Then logic deductions are used from these assumptions to obtain the correct (valid) model. Empiricism requires every assumption and outcome to be empirically verified. This approach is the opposite of rationalism. Positive Economics requires only that the model be able to predict the future and is not concerned with its assumptions or mechanisms (structure).

Multistage Validation: Naylor and Finger [13] proposed combining the three historical methods of Rationalism, Empiricism, and Positive Economics into a multistage process of validation. This validation method consists of (1) developing the model's assumptions on theory, observations, general knowledge, and intuition; (2) validating the model's assumptions where possible by empirically testing them; and (3) comparing (testing) the input-output relationships of the model to the real system.

Internal Validity: Several replications (runs) of a stochastic model are made to determine the amount of stochastic variability in the model. A high amount of variability may cause the model's results to be questionable and therefore should be reduced, if possible.

Parameter Variability - Sensitivity Analysis: This validation technique consists of changing the input and internal parameters of a model to determine the effect upon the model and its output. The same relationships should occur in the model as in the real system. Those parameters that are sensitive, i.e., cause sufficient changes in the model's behavior, should be made sufficiently accurate prior to using the model. (This may require iterations in model development.)

Comparison to Other Models: Various results, e.g., output, of the simulation model being validated are compared to results of other models. (For example, simple cases of the simulation model may be able to be compared to known analytical results.)

Historical Data Validation: If historical data exist (or if data is collected on a system), part of the data is used to build the model and the remaining data is used to determine (test) if the model behaves the same as the system does.

Predictive Validation: The model is used to predict (forecast) the system behavior and tests are made to determine if the system behavior and the model's forecast are the same.

Event Validity: The "events" of occurances of the simulation model are compared to those of real system to determine if they are the same. (Examples of events might be the number of deaths in a given fire department simulation or number of fires having a given amount of fire damage.)

Live Graphics: Computer output graphics of the model's operational behavior are evaluated. (An example is the flow of traffic through an intersection in a traffic simulator.)

Submodel Testing: Submodels of the simulation model are validated using the techniques described here.

USEAGE OF VALIDATION TECHNIQUES

In the preceding section, various validation techniques were described. One can apply these techniques either subjectively, such as using graphical approaches, or objectively by using statistical tests. In either case, it is necessary to have data from the real system if a high degree of confidence is to be obtained in a model. In validating a model, usually several of the validation techniques are selected and used by the analyst. As stated above, there is no algorithm or procedure for selecting the techniques to use in a given model validation. Although there is no fixed procedure, face validation is almost always used initially to determine if the model is reasonable. Also statistical tests should be used where feasible and economical.

Because there is no procedure or algorithm for selecting and using the validation techniques presented above, this author suggests the following steps as a minimum be performed in validation: (1) an agreement be made prior to building the model between the model's sponsor and the modelling team as to the validation approach to be used, (2) theoretical validity be performed in developing the flowchart model during each model iteration, (3) operational validaty be performed after the simulator has been verified, and (4) validation be discussed in the model documentation. The validation approach to be determined in step 1 does not refer to which specific validation techniques are to be used but instead to determining (i) whether the modelling team is to determine if the model is valid, whether a specific set of system data must be used in validation, whether the model's sponsors are going to use an independent evaluation team in addition to the validation performed by the modelling team, etc., (ii) the accuracies desired between the model and the system for given sets of conditions, e.g., whether certain statistical tests must be satisfied, face validation is to be used to determine if the model obtains the necessary "accuracy", etc., and (iii) the type and amount of information regarding validation that is to be in the model's documentation, e.g., is all the data used in validation to be included or just the results of the statistical tests, etc.

In general, at least the following are of concern in validation: (1) Are the assumptions underlying the model correct? (2) Are the parameters, the statistical distributions, and other data used in the model correct? (3) Are the model's input-output transformations correct? To answer these questions, data needs to be collected on the real system and analyzed, preferably statistically, for at least two different operating levels. In collecting and analyzing the system data, one must be concerned with whether the data is independent, correlated, stationary, or nonstationary. Depending on the behavior of the data and the method of collecting it, the appropriate statistical techniques must be used. To answer the first two questions, estimates must be made of parameters, distributions, etc. To answer the last question, comparisons must be made of the model's and system's outputs for given inputs. The latter might be concerned with just means, variances, and distributions or might be concerned with time series.

The statistical techniques used might be straight forward, such as estimating a mean or comparing two means, or it might require the state-of-the-art tests such as comparing two non-stationary time series. The statistical techniques used in validation can include a vast majority of statistics, e.g., classical statistics, non-parametric statistics, and time series analysis. The statistical tests used are described in statistic books as well as in articles on validation and simulation texts [7, 12, 14, 17, 18, 20].

If statistical tests cannot be used, the data from the system and the model should be examined using other techniques such as graphs, histograms, etc. If system data is not available, then "guestimates" must be used for the model; however, the model's output, etc. should still be examined as thoroughly as possible.

Various results using statistical tests for validation have been reported. Chandra and Sargent [2] have reported on a simulation study of a rework area of a production system in which the simulator was validated by determining that the 95% confidence intervals of the mean production rates and the utilizations of the repairmen from the simulator contained the means from the system. Teorey [20] reports on a simulation model of a computer system that he was able to validate using statistic tests to compare means, variances, and distributions of several model and system variables using statistical tests requiring normal distribution assumptions. Kheir and Homes [12] report on validating simulation models of missile systems using Theil's inequality coefficients. Stafford [19] reports on a simulation model of a multifacility out-patient clinic that he was able to validate using statistical tests and the validation criteria specified by Schellenberger [16]. Anderson and Sargent [1], however, report on a simulation study of a computer system that they were not able to use statistical tests in validation because of the large amount of variability occurring in the system and the model. Instead they made graphical comparisons of various model and system variables. In [21], Trehan and Sargent report on a simulation model used to aid in the design of a large computer/manual microfiche information storage and retrieval system for the Air Force where system data did not exist and thus "guestimates" were used. They were able to answer the designers questions and, in addition, pointed out a problem area that the designers were unaware of.

SUMMARY

This paper described what is meant by validation and the validation techniques used to obtain confidence in simulation models. It was stated that validation techniques are problem dependent, that it is desirable to have system data, and that it is desirable, if possible, to use statistical tests in applying validation techniques. Furthermore, it was pointed out that both the modeller and the user must have confidence in the model. One, however, should be cautioned on being too stringent in comparing the model and the system. As Fishman [6] states, it is his feeling that too much emphasis can be placed on model versus system statistical exactness instead of placing validation in the decision making context, i.e., does the use of the model lead to the same action by a decision maker as would be made by using knowledge from the real system.

BIBLIOGRAPHY

[1] Anderson, H.A. and R.G. Sargent, "Investigation into Scheduling for an Interactive Computer System," IBM Journal of Res. & Dev., Vol. 18, No. 2, March 1974.

[2] Chandra, M.J. and R.G. Sargent, "An Investigation of the Effect of Product Quality on Rework Manpower During Production Via Simulation," Proceedings of 1978 Winter Simulation Conference, Edited by Highland, et al.

[3] Conway, R.W., "Some Tactical Problems in Digital Simulations," Management Science, Vol. 10, No. 1, October 1963, pp. 44-61.

[4] Emshoff, J.R. and R.L. Sisson, Design and Use of Computer Simulation Models, MacMillan, 1970, pp. 204-206.

[5] Herman, C.F., "Validation Problems in Games and Simulations With Special Reference to Models of International Politics," Behavior Science, Vol. 12, 1967, pp. 216-231.

[6] Fishman, G.S., "On Validation of Simulation Models," Proceedings AFIPS 1973 National Computer Conference, Vol. 42, pp. 51.

[7] Fishman, G.S., Concepts and Methods in Discrete Event Digital Simulation, Wiley, 1973.

[8] Fishman, G.S. and P.J. Kiviat, "The Statistics of Discrete Event Simulation," Simulation, Vol. 10, 1968, pp. 185 (same as in Rand Report RM-5387-PR).

[9] Gass, S.I., "Evaluation of Complex Models," Computers and Operations Research, Vol. 4, pp. 27-35, March 1977.

[10] Gass, S.I., "A Procedure for the Evaluation of Complex Models," Proceedings of the First International Conference on Mathematical Modeling, August 24-September 1, 1977.

[11] Guidelines for Model Evaluation (Exposure Draft), U.S. General Accounting Office, PAD-79-17, January 1979.

[12] Kheir, N.A. and W.M. Homes, "On Validating Simulation Models of Missile Systems," Simulation, Vol. 30, No. 4, April 1978.

[13] Naylor, T.H. and J.M. Finger, "Verification of Computer Simulations Models," Management Science, Vol. 14, No. 2, October 1967, pp. 1B92-1B101.

[14] Sargent, R.G., "Statistical Analysis of Simulation Output Data," ACM Proceedings of the Symposium on the Simulation of Computer Systems IV, 1976, and in Simuletter, Vol. 8, No. 3, April 1977.

[15] Sargent, R.G., "Validation of Discrete Event Simulations Models," Proceedings of the NSF Workshop on Validation of Mathematical Models in Energy Related Research and Development, June 21-23, 1978, Fort Worth, Texas.

[16] Schellenberger, R.E., "Criteria for Assessing Model Validity for Managerial Purposes," Decisions Science, Vol. 5, No. 4, October 1974, pp. 644-653.

[17] Shannon, R.E., Systems Simulation: The Art and Science, McGraw-Hill, 1956.

[18] Siegel, S., Nonparametric Statistics for Behavior Sciences, McGraw-Hill, 1956.

[19] Stafford, E.F., Jr., "Technical Validation of a Simulation Model," Working Paper 76:7, College of Business Administration, University of Oklahoma, Norman, Oklahoma.

[20] Teorey, Toby, Jr., "Validation Criteria for Computer Systems Simulation," Proceedings of the Symposium on the Simulation of Computer Systems III, 1975.

[21] Trehan, V. and R.G. Sargent, "A Simulation Model of the Air Force Military Center (AFMPC) Initial Operating Capability (IOC) Microform System," Proceedings of the 1976 Summer Computer Simulation Conference.

[22] Van Horn, R.L., "Validation of Simulations Results," Management Science, Vol. 17, No. 5, January 1971, pp. 247-258.

[23] Zeigler, B.P., Theory of Modelling and Simulation, Wiley, 1976.